CalEDAS Database Status and Plans   *June 15, 2004*

Background

Development of a database is a fundamental first step for any coordinated bioassessment program (see Barbour and Hill 2002, http://www.swrcb.ca.gov/swamp/reports.html). The US EPA and TetraTech, Inc. created and developed the EDAS (Environmental Data Analysis System) database to manage various bioassessment programs throughout the country. In 1999, the Department of Fish and Game's Aquatic Bioassessment Laboratory (ABL) began modification of the EDAS database to accommodate special needs of California datasets. Over the past four years, we have made extensive modifications/ improvements to the original design to create CalEDAS, an MS Access database that captures all aspects of our bioassessment program. We use this database to manage samples and sampling location information, associated physical habitat and chemistry data, track sample processing, organize our taxonomic information and support our increasingly sophisticated reporting needs. *Note: In its current form, CalEDAS bears little resemblance to the original EDAS design; the database is likely to be renamed before the next major revision is complete.*

Availability of Database

CalEDAS was created to meet ABL data storage and reporting needs, but its design can be modified to accommodate the needs of other programs (e.g. SNARL, UCD-ATL, citizen monitoring, private labs, etc). By late 2005, this database will serve as the standard bioassessment database for the State Water Quality Control Board's bioassessment data. The current version of CalEDAS has the structure to handle benthic macroinvertebrate data and we are currently working on the integration of physical habitat and chemistry data. However, although the current version of CalEDAS has most of the structure needed to handle bioassessment data, it is utilitarian but not very user-friendly.

We will provide copies of the database to anyone at no cost, but we do not have the resources to provide any user support at this time. As part of the ongoing effort to integrate CalEDAS with the current SWAMP database, we will be developing user tools (e.g. user's manuals, data dictionaries) to support the database, but these will be managed by the State Water Quality Control Board. Thus, future versions of the database should be much more user-friendly, but the ABL has no future plans to provide user support.

We strongly recommend that anyone interested in using CalEDAS in its current version have access to someone with database expertise.

Overview of Future Plans for CalEDAS

The ABL is currently completing work on its internal version of CalEDAS. The CalEDAS database has potential to fulfill two essential roles for the main bioassessment labs that contribute data to SWAMP:

1. CalEDAS can be used as a stand alone tool for managing and reporting each group's own biotic and associated environmental data internally (see attached Figure 2)
2. CalEDAS can serve as a mechanism for transferring individual program data to the SWAMP database (see attached Figure 1)

The ABL is coordinating closely with MLML and DWR to make CalEDAS compatible with the ongoing SWAMP database effort, but there is a growing need to make this happen faster than ABL resources can manage. We are seeking to accelerate the completion of the CalEDAS components so that it can be deployed to the Sierra Nevada Aquatic Research Laboratory (SNARL), SWAMP and any other interested parties as quickly as possible. We have recently received funding from the State Water Quality Control Board's SWAMP program for a database programmer to work with ABL staff and SNARL to complete the basic CalEDAS design and to work with DWR staff and MLML staff to ensure data compatibility. We anticipate that this level of funding will be enough to establish a functioning mechanism; the need for future funding for support and refinement of this mechanism can be revisited in a year.

The proposed funding would be managed by Rusty Fairey and the SWAMP database management group at Moss Landing Marine Laboratories (MLML), who would use the money to pay for ABL and DWR programmer time. The database design work and testing would be overseen by ABL staff. The ABL has invested approximately $200,000 to date in staff scientist time and database consulting fees to design and beta test this database. ABL will continue to donate its staff time and fund consulting time with our database designer to help us coordinate the proposed work.

CalEDAS Remaining Tasks (work to be addressed under this proposal **in bold)**

I.   Completing basic CalEDAS design
      a. Finish Main Reporting module
          i. Beta testing of main reporting module
         ii. Add new standard reports
      b. Enhance Samples and Stations data forms

    c. Standardize criteria and rules for taxonomic data with ITIS/ STORET

    d. Complete QA/QC module and assimilate into CalEDAS

    e. Develop structure for entering and reporting physical habitat data and ambient chemistry data
- i. CSBP, SNARL, EMAP-multi-habitat, EMAP-targeted riffle
- ii. Ensure compatibility with existing SWAMP chemistry and physical habitat standards

    f. Develop structure for *Fish* and *Algae* samples

    g. Data Loading Module
- i. Complete development of MS Excel-based data loading conversion utility for getting existing data (SNARL/UCD) into CalEDAS and integrate module into CalEDAS
- ii. Needs to accommodate other labs needs

    h. Integrating probabilistic site selection data

II. CalEDAS Data Export to DWR/ SWAMP
    a. Work out bugs in partial replica model

III. Accommodating alternate data sources (SNARL, UCD, Region 5, others)
    a. Physical habitat
    b. Chemistry
    c. Site/ sample information
    d. Reporting needs

IV. Enhance Data Validation
    a. Application of constraints
    b. Cleanup of historical data

V. Translation of reporting tools to BDAT environment
    a. Standard reports: taxa and metrics
    b. Specialized reporting:
- i. Monte Carlo subsampling
- ii. Distinct/ Non-distinct coding (standardizing "ambiguous" taxa)
- iii. Reporting metrics at different levels of taxonomic effort
- iv. Calculation of IBI scores, O/E scores
- v. QA/QC reporting and analysis

VI. Developing User Support
    a. Database Documentation
- i. Administrative documentation
- ii. Programmer's documentation
- **iii. Data dictionary**
- **iv. User's manual**

    b. Set up arrangements for end user support of CalEDAS and BDAT reporting
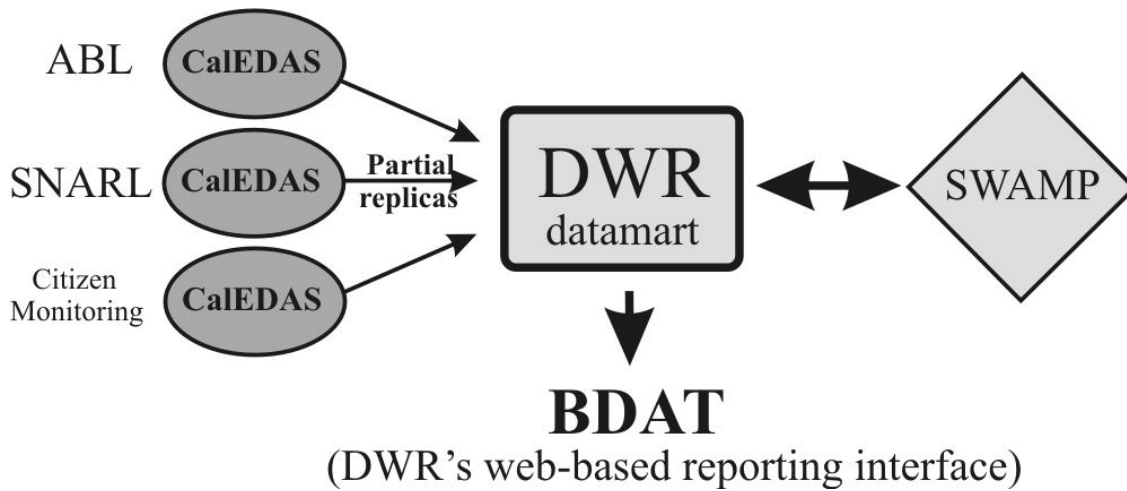
VII. LOTS of BETA TESTING

Figure 1. Proposed model for getting bioassessment data into the SWAMP database. Note: The "partial replica" design allows CalEDAS users to control which data are shared with DWR/SWAMP. The initial goal of this proposal is to get the structure to work for ABL and SNARL, but the basic design should be easily adapted for citizen monitors and other labs.
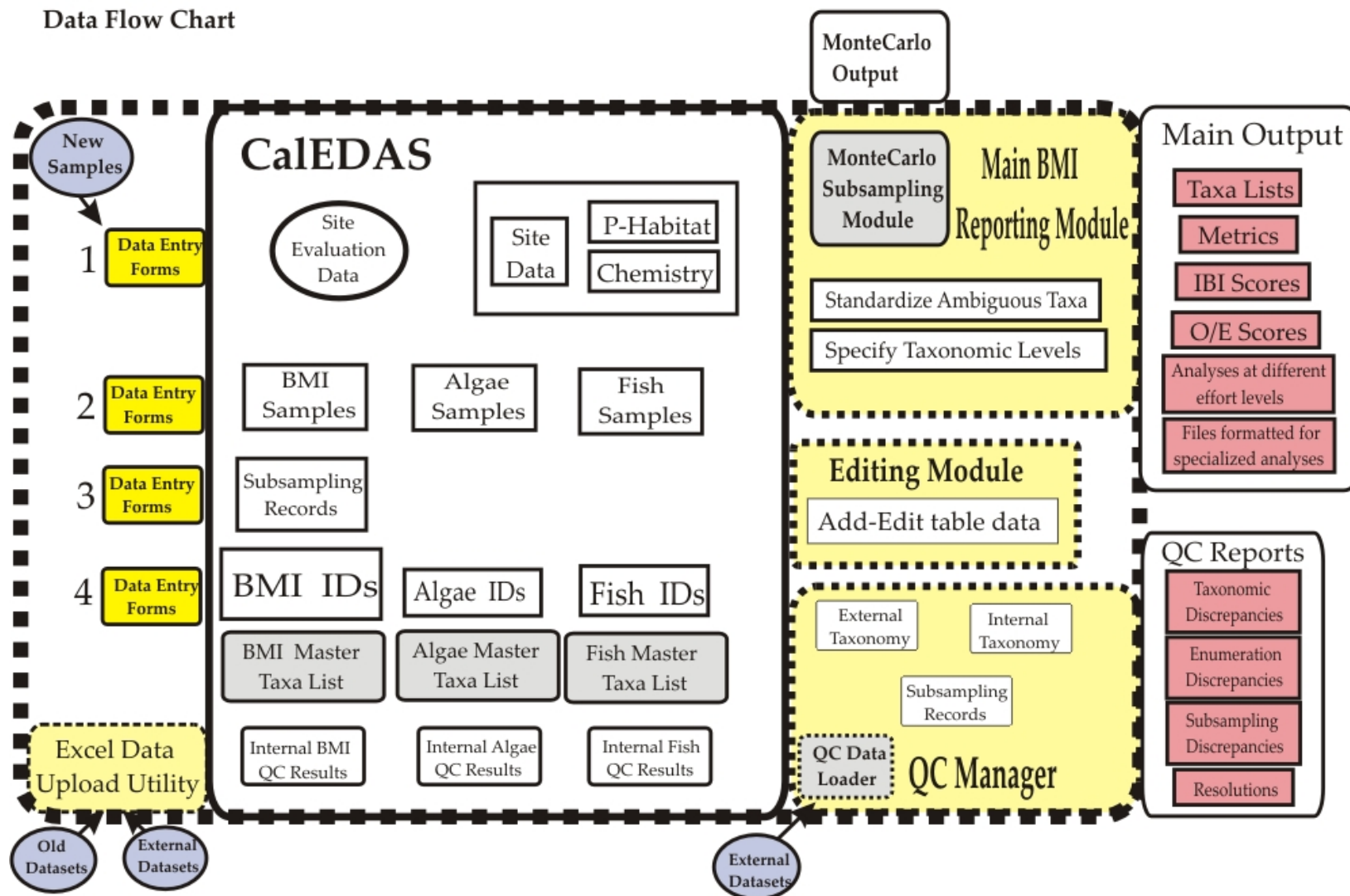
Figure 2. Partial data model for proposed version of CalEDAS (03/31/2004)